

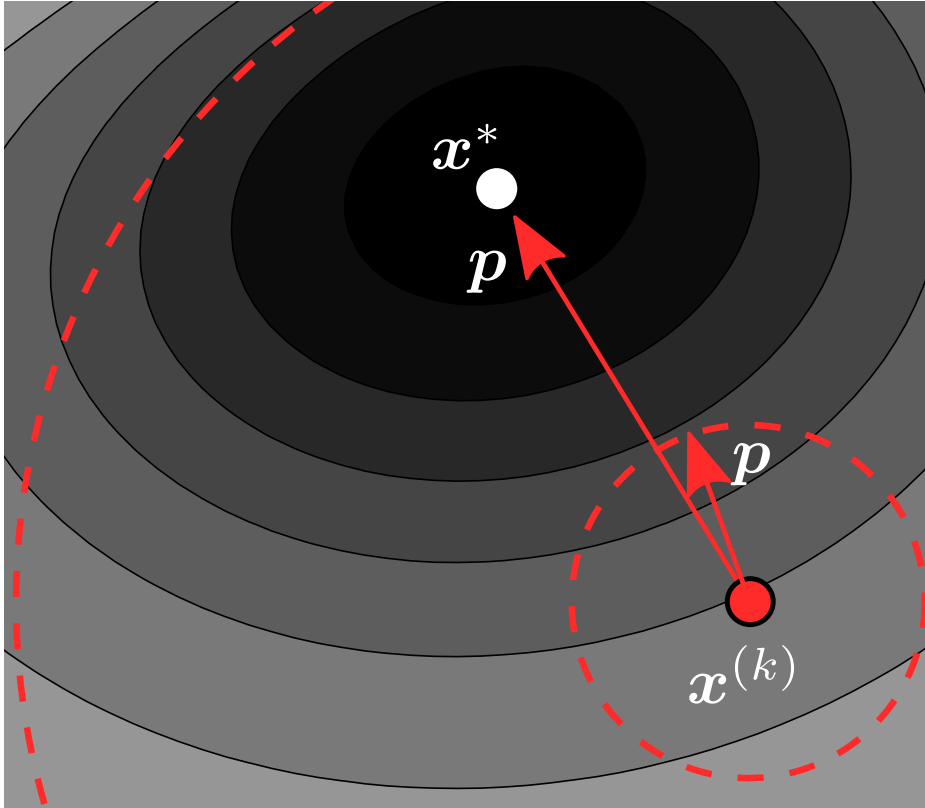
TRUST REGION METHODS

David F. Gleich

April 3, 2023

The material here is from Chapter 4 in Nocedal and Wright.

1 PICTORIAL DESCRIPTION OF TRUST REGION METHODS



2 THE TRUST REGION MODEL

Taylor's series gives us:

$$f(\mathbf{x}_k + \mathbf{p}) = f_k + \mathbf{g}_k^T \mathbf{p} + \frac{1}{2} \mathbf{p}^T \tilde{\mathbf{H}} \mathbf{p}$$

where $\tilde{\mathbf{H}} = \mathbf{H}(\mathbf{x}_k + t\mathbf{p})$ for some $0 \leq t \leq 1$. The ideal with the trust region method is to use this model to find a direction \mathbf{p} within a “trusted region”. That is, we know that this approximation is only valid in a small area, so let's pick something to approximate $\mathbf{B}_k \approx \tilde{\mathbf{H}}$ and then find:

$$\begin{aligned} & \underset{\mathbf{p}}{\text{minimize}} && f_k + \mathbf{g}_k^T \mathbf{p} + \frac{1}{2} \mathbf{p}^T \mathbf{B}_k \mathbf{p} \\ & \text{subject to} && \|\mathbf{p}\| \leq \Delta. \end{aligned}$$

3 FIRST OBSERVATIONS

If Δ is sufficiently large and $\mathbf{B}_k > 0$, then this will correspond to

$$\underset{\mathbf{p}}{\text{minimize}} \quad f_k + \mathbf{g}_k^T \mathbf{p} + \frac{1}{2} \mathbf{p}^T \mathbf{B}_k \mathbf{p} .$$

The minimizer of this is $\mathbf{B}_k \mathbf{p} = -\mathbf{g}$, the *Newton step*. In trust region lingo, this is called *the full step*.

If Δ is sufficiently small, then what happens? We need a bit more info to tell us.

We'll work through this in a little bit, but the theory of constrained optimization tells us that the solution of this problem is:

$$(\mathbf{B}_k + \lambda \mathbf{I}) \mathbf{p} = -\mathbf{g}.$$

where $\lambda \Delta - \lambda \|\mathbf{p}\| = 0$, and $\mathbf{B}_k + \lambda \mathbf{I}$ is positive semi-definite.

CLASS QUESTION What happens to $\|\mathbf{p}\|$ as λ gets larger?

This is related to Tikhonov regularization! Recall that Tikhonov regularization is:

$$\text{minimize} \quad \|\mathbf{b} - \mathbf{A}\mathbf{x}\|^2 + \lambda \|\mathbf{x}\|^2.$$

We could solve this system by solving:

$$(\mathbf{A}^T \mathbf{A} + \lambda \mathbf{I}) \mathbf{x} = \mathbf{A}^T \mathbf{b}.$$

What did we see for $\lambda \rightarrow 0$ and $\lambda \rightarrow \infty$?

QUIZ What is the smallest value that λ could possibly be for a given \mathbf{B}_k ? (debugging code motivation)

SOLUTION Since $\mathbf{B}_k + \lambda \mathbf{I}$ must be positive semi-definite, all its eigenvalues must be larger than 0. Let the eigenvalues of \mathbf{B}_k be $\lambda_1 \leq \dots \leq \lambda_n$. Then the eigenvalues of $\mathbf{B}_k + \lambda \mathbf{I}$ are $\lambda_1 + \lambda \leq \dots \leq \lambda_n + \lambda$; we need $\lambda_1 + \lambda \geq 0$, so $\lambda \geq -\lambda_1$. Thus, solutions of the problem must have $\lambda \in [-\lambda_1, \infty)$.

If $\lambda = 0$, then we get the Newton step! So what happens as $\lambda \rightarrow \infty$? Obviously, $\mathbf{p} \rightarrow 0$.

QUIZ But what direction does it go to (if any) as $\lambda \rightarrow \infty$? Argue or prove. For large enough λ , the matrix $\mathbf{B}_k + \lambda \mathbf{I} \approx \lambda \mathbf{I}$, in which case, $\mathbf{p} \approx -\frac{1}{\lambda} \mathbf{g}$. This is the steepest descent direction.

Consequently, the trust region method “interpolates” between Newton’s method (large Δ) and steepest descent (small Δ)

4 A TRUST REGION METHOD

Given a maximum trust region size $\Delta_{max} > 0$

Given an initial region $0 < \Delta_0 < \Delta_{max}$

Given a parameter η

while not done

 Compute the **next** step \mathbf{p}_k by solving

$$\text{minimize}_{\mathbf{p}} \quad f_k + \mathbf{g}_k^T \mathbf{p} + \frac{1}{2} \mathbf{p}^T \mathbf{B}_k \mathbf{p} \quad \text{subject to } \|\mathbf{p}\| \leq \Delta_k.$$

 Check how well your model matches the function by

$$\text{evaluating } \rho_k = \frac{f(\mathbf{x}_k) - f(\mathbf{x}_{k+1})}{m_k(0) - m_k(\mathbf{p}_k)}$$

 # Update the trust region

 If $\rho_k < 1/4$, then **set** $\Delta_{k+1} = \Delta/4$

 If $\rho_k > 3/4$ **and** $\|\mathbf{p}_k\| = \Delta_k$, **set** $\Delta_{k+1} = \min(2\Delta, \Delta_{max})$

 If $\rho_k > \eta$, $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{p}_k$

 Else, $\mathbf{x}_{k+1} = \mathbf{x}_k$

5 SOLVING THE SUBPROBLEM

See solve_trust_regions_subproblem.m. It’s nasty. It involves trying to find a root of $\|\mathbf{p}(\lambda)\| - \Delta$. Each iteration involves solving a linear system!

6 MAKING IT PRACTICAL

We need something faster! For line search methods, we didn’t need an exact line search. Instead, we just needed to make enough progress. Something similar will hold true here.